Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

DATA FARMING AROUND THE WORLD OVERVIEW

Gary E. Horne

SEED Center for Data Farming
Naval Postgraduate School
Monterey, California, USA

Klaus-Peter Schwierz

System Design Center Germany
European Aeronautic Defense and Space Company
Friedrichshafen, Germany

ABSTRACT

Data Farming combines the rapid prototyping capability inherent in certain simulation models with the exploratory power of high performance computing to rapidly generate insight into questions. The Data Farming process focuses on a more complete landscape of possible system responses, rather than attempting to pinpoint an answer. Data Farming allows decision makers to more fully understand the landscape of possibilities and also consider outliers that may be discovered. Over the past decade, an international community has formed around these ideas. In 2008, International Data Farming Workshop 16 took place in Monterey, California, USA and workshop number 17 was held in Garmisch Partenkirchen, Germany. In addition to a summary of these two workshops, this paper will present an overview of the process that has developed to include the development of both methods and applications in the International Data Farming Community.

1 INTRODUCTION

Data Farming was first introduced to the Winter Simulation community at the 1999 Winter Simulation Conference in Phoenix (Horne 1999). The ideas behind Data Farming had been initially developed much earlier, but were introduced by Dr. Horne to the defense community in 1997 (Horne 1997) in concert with the combination of agent-based models with high performance computing that was the start of Project Albert.

Project Albert was a congressionally funded modeling and simulation initiative of the United States Marine Corps (USMC) motivated by the fact that complex adaptive systems are pervasive in USMC operations. The philosophy of Project Albert was to pair simple, efficient, abstract models with high performance computing to explore large design spaces. When these models and high performance computing are combined with efficient experimental designs developed in work pioneered at NPS (e.g. see Sanchez and Lucas 2002), a huge sample space can be explored very rapidly. And when rapid prototyping capabilities and collaborative environments are introduced into the Data Farming process, progress on questions, even long-standing and difficult questions involving many interacting variables, is possible.

Project Albert used what are referred to as agent-based distillation models. These are a type of computer simulation which attempts to model the critical factors of interest in combat without explicitly modeling all of the physical details. Some of the models used in Project Albert were MANA, PAX, and Pythagoras, all agent-based models, although the methods developed can be applied using any type of simulation model. These models continue to be developed and recent updates are described in Lauren (2007), Lampe (2007), and Henscheid (2007) respectively. But agent-based models are small and abstract and can easily be run many times to test a variety of parameter values and get an idea of the landscape of possibilities. The term distillation is added, because the intent is to distill the question at hand down into as simple a representation as possible. Also, models used in Project Albert were specifically developed and used because the capability to rapidly prototype scenarios is very important in the process.

Although Project Albert was a US sponsored effort, it had a strong spirit of international collaboration which made possible a great deal of cooperative effort among researchers around the world. For example, the three models mentioned in the paragraph above were developed in New Zealand, Germany, and the US respectively. And the fact that Project Albert was question-based also allowed practitioners from around the world to rally around the developing Data Farming methodologies because of the impact upon their shared application interests.

Because many of the questions of interest have wide applicability, the work teams at international workshops typically consist of representatives from two or three and sometimes up to six different countries and the workshops overall have usually consisted of about 8 to 10 teams. The first international workshop was organized in 1999 and 12 workshops were held under the auspices of Project Albert which ended in September 2006. Since that time the international Data Farming community has continued coopera-

978-1-4244-2708-6/08/$25.00 ©2008 IEEE

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Data Farming Around the World Overview	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,SEED Center for Data Farming,Monterey,CA,93943**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the Proceedings of the 2008 Winter Simulation Conference, 7-10 Dec, Miami, FL

14. ABSTRACT
Data Farming combines the rapid prototyping capability inherent in certain simulation models with the exploratory power of high performance computing to rapidly generate insight into questions. The Data Farming process focuses on a more complete landscape of possible system responses rather than attempting to pinpoint an answer. Data Farming allows decision makers to more fully understand the landscape of possibilities and also consider outliers that may be discovered. Over the past decade, an international community has formed around these ideas. In 2008, International Data Farming Workshop 16 took place in Monterey, California, USA and workshop number 17 was held in Garmisch Partenkirchen, Germany. In addition to a summary of these two workshops, this paper will present an overview of the process that has developed to include the development of both methods and applications in the International Data Farming Community.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

tive efforts and the most recent workshops, the 16th and 17th workshops, were held in 2008. Our presentation here will highlight these most recent workshops, but details on the other workshops and other Data Farming efforts can be found at http://harvest.nps.edu, the website of the SEED (Simulation Experiments and Efficient Designs) Center for Data Farming at the United States Naval Postgraduate School.

2 DATA FARMING

2.1 Overview

Data Farming combines the rapid prototyping of agent-based distillations with the exploratory power of high performance computing to rapidly generate insight into military questions. Data Farming focuses on a more complete landscape of possible system responses, rather than attempting to pinpoint an answer. This "big picture" solution landscape is an invaluable aid to the decision maker in light of the complex nature of the modern battlespace. And while there is no such thing as an optimal decision in a system where the enemy has a role, Data Farming allows the decision maker to more fully understand the landscape of possibilities and thereby make more informed decisions. Data Farming also allows for the discovery of outliers that may lead to findings that allow decision makers to no longer be surprised by surprise.

The simulations that defense analysts use are often large and complex. Also, even the smaller more abstract agent-based distillations referred to above can have many parameters that are potentially significant and that could take on many values. And response surfaces can be highly non-linear. Thus, even with high performance computing and the small models used in Data Farming, gridded designs where every value is simulated are unwieldy.

Thus, using efficient experimental designs is essential and work in this area has been performed at the Naval Postgraduate School (NPS) in Monterey, California and NPS researchers have collaborated with others worldwide as well (see Kleijman, Sanchez, Lucas, and Cioppa 2005). Data Farming continues to evolve from initial Project Albert efforts (Hoffman and Horne 1998) to the work documented in the latest edition of the *Scythe* (Horne and Meyer 2008b). This publication contains the proceedings of the International Data Farming Workshops that have taken place since Project Albert ended and is put out by the SEED Center for Data Farming at NPS.

2.2 Question-Based

Over the past few years several articles have captured the fundamentals of Data Farming (e.g. Horne and Meyer 2005 and Lawler 2005), but the key is the question at hand. At the Naval Postgraduate School over 60 theses, many by international students, have been completed which have used data farming over the past decade and as we shall present in the next section, over 100 international work teams have formed around questions at International Data Farming Workshops.

These types of questions can never have precisely defined initial conditions and a complete set of algorithms that describe the system being considered. These questions address open systems that defy prediction. Data Farming is used to provide insight that can be used by decision-makers. To accomplish this formidable task, Data Farming relies upon two basic ideas:

1. use high performance computing (HPC) to execute models many times over varied initial conditions to gain understanding of the possible outliers, trends, and distribution of results, and

2. develop models, called distillations, that are focused to specifically address the question.

Data Farming, by providing the ability to process large parameter spaces, makes possible the discovery of surprises (both positive and negative) and potential options.

2.3 Iterative Process

Data Farming is an iterative team process (Horne and Meyer 2004b). Figure 1 presents the Data Farming process as a set of imbedded loops. This process normally require input and participation by subject matter experts, modelers, analysts, and decision-makers.

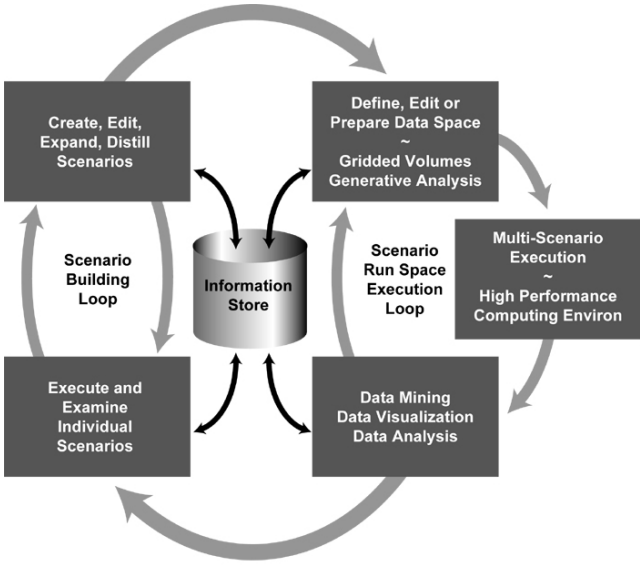

Figure 1: Data Farming Iterative Process

The "Scenario Creation" loop shown on the left side of the figure involves developing and honing a model that adequately represents the system that addresses the ques-

tion being asked by the decision-maker. This is an iterative process that often requires honing the question as well.

The "Scenario Run Space Execution" loop shown in Figure 1 is entered once the *basecase* of the scenario is complete. In this loop the team defines a *study* which determines which scenario input parameters should be examined and what processes should be used to vary them. Here the team is exploring the possible variations (or *excursions* of the basecase) in the initial conditions of the scenario. Specifically those parameters that address the question being posed are considered.

The defined study is used to guide the execution of many runs of the model in the HPC environment. Each run produces output which is collected by the Data Farming system and provided as output to analysis capabilities. After analysis of the results, the team (or an algorithm) may decide to adjust or produce a new study or adjust the model to more adequately address the question. This process continues until insight related to the decision-maker's question has been gained.

3 WORLD WIDE PARTNERS

The first international workshop took place in 1999. Some of the cities and towns where the Project Albert and International Data Farming Workshops took place are world-reknowned and some perhaps are not. They include Kihei, Auckland, Cairns, Uberlingen, Monterey, Quantico, Singapore, Wellington, Stockholm, Honolulu, Boppard, Den Hague, and Garmisch Partenkirchen. But the actual workshops, in fact celebrations of sorts, are merely the culmination of a great deal of work that takes place in between them. Collaboration and contributions to the overall advancement of Data Farming takes place in the development of simulation models, scenarios within the models, and computer clusters to run the models audacious numbers of times.

But the real work is in making progress on important questions and the real secret is the combination of military subject matter experts and highly knowledgeable and multi-disciplinary scientists. This special mix of personnel has been the hallmark of the international workshops. It has been a dynamic combination to have Data Farming work teams headed up by a person who really knows and cares about the question (e.g. a military officer who knows that the answers may have an impact on both mission success and lowering casualties) and supported by men and women with technical prowess who can leverage the tools available.

Countries represented throughout the decade of workshops include, Australia, Canada, Germany, Mexico, New Zealand, Norway, Portugal, Singapore, South Korea,, Sweden, Turkey, the United Kingdom, and the United States. But each nation involved in Data Farming has its own story regarding their contributions and what they have received from their participation.

Here we will now provide some details regarding how German involvement increased from their participation in the early workshops when the methodologies were in the beginning stages of development. After Project Albert started the emphasis on using Data Farming tools within combat situations and the simulation tools were developed to represent these situations, the German Delegation carried in questions regarding peace support operations. Human factors modeling and the influences of intangibles are becoming more and more essential in this question area. To simulate the non-attrition based parts in peace support operations the model PAX (after the Roman goddess of peace) was developed in Germany and released to the International Data Farming Community. The contributions led to a broad acceptance of Data Farming in the German modeling and simulation community.

The Project Albert mission was clearly to develop the methodology of Data Farming in collaborative environments. All German applications had the clear goal: Not to replace the classical Modeling and Simulation tools by new ones but to apply both methods in an "operational synthesis" (see Brandstein 1999). The application of complex adaptive systems theory with the modeling following the agent based paradigm had the goal to explore the wide field of non linearity, of co-evolution and intangibles. Results were a continuum of solutions in the sense of optimization theory with the relevant tools for a statistical experimental design and the semi-automated evaluation techniques directing the user to unknown effects, or "surprises" and interrelations in the analysis of a variety of possible progressions.

The international community drove the German mission through other methodological developments such as Nearly Orthogonal Latin Hypercube experimental designs (e.g. see Cioppa 2002), model developments (MANA, Pythagoras, PAX, etc.), application of agent based development environments (NetLogo, REPAST, MASON, etc.) and through evaluation and analysis tools of many types (e.g. see Upton 2004). In the international workshops the availability of the experts and the free and open information sharing led to a big success in Germany in the application of the tools (Schweirz 2008).

Other Data Farming efforts around the world are documented in a variety of places. The beginnings of development in the United States is documented in *Maneuver Warfare Science 1998* (Hoffman and Horne 1998). And additional volumes of Maneuver Warfare Science from 2001, 2002, and 2003 contain contributions from the US as well as Sweden, New Zealand, Australia, and Singapore. Also, the book *It's Alive* (Meyer and Davis 2003) contains a chapter describing some of the initial USMC efforts. Many presentations involving Data Farming have also been made at INFORMS meetings (e.g. Horne and Meyer

2004a) and MORS Symposia (e.g. McDonald 2008) over the past decade and Winter Simulation Sessions on Data Farming were held in 2004, 2005, and this year. Finally, the *Scythe* is a regular publication from the SEED Center for Data Farming that documents workshop proceedings.

4 INTERNATIONAL WORKSHOP TOPIC SUMMARY

Through international workshop 16 we have had over 100 work teams in a variety of areas. Of course some of those work teams have continued from workshop to workshop. For example the combat identification team which has had representation over the years from 6 different countries started at workshop 12 and continues to the present. These 100+ work teams do fall into areas, or themes, which include: Joint and Combined Operations (e.g. C4ISR Operations, Network Centric Warfare, Networked Fires, and Future Combat Missions), Urban Operations, Combat Support (e.g. UAV Operations, Robotics, Logistics, and Combat ID), Peace Support Operations, the Global War on Terrorism, Homeland Defense, and Disaster Relief. Other work teams have looked into continuous support of modeling such as efficient designs, new models, model improvements, automated red teaming, and automated coevolution. In the next section we will present some of the specifics of recent workshops.

5 INTERNATIONAL WORKSHOPS IN 2008

Two Workshops were held in 2008. International Data Farming Workshop (IDFW) 16 was held in Monterey, California, USA from 13 to 18 April and IDFW 17 was held in Garmisch Partenkirchen, Germany 21-26 September. The latter occurred too late to be include in this summary, but the 11 teams from IDFW 16 are listed below to give a flavor of the breadth of topics explored at an IDFW (Meyer and Horne 2008b).

Team 1 used Pythagoras to explore the contribution of small unmanned ground vehicles to small unit combat effectiveness. The team developed a building clearing scenario and examined different vehicle capabilities such as speed, sensor range, and vulnerability.

Team 2 built on the work of a completed NPS thesis which examined questions regarding the new Littoral Combat Ship using MANA. The team illustrated the power of Data Farming by conducting over 40,000 replications to help understand the implications of a variety of possible red tactics.

Team 3 was an internationally co-led team which used both the PAX and MANA models and applied Automated Red Teaming to investigate different aspects of the same problem involving peace support operations. The scenario used in this team's Data Farming was based on a crowd control situation in a stabilization operation.

Team 4 used the opportunity to participate in IDFW 16 to begin an effort using the agent-based sensor effector model (ABSEM) recently developed in Germany. They presented the main ideas of the ABSEM to learn from the available expertise and they plan to data farm a prototype in the future.

Team 5 used the Logistics Battle Command model and experimental design techniques to assess the impact that Soldier level network enabled capabilities have on cargo operations at a truck terminal node within a sustainment base supporting a joint force.

Team 6 was led by personnel from the Joint Test and Evaluation Methodology program. This team applied design of experiments and Data Farming using MANA for developing evaluation strategies for testing in a joint environment.

Team 7 not only won the best poster competition, but used Data Farming to explore parameters and assumptions using the Total Life Cycle Management-Assessment tool on a Marine Light Armored Vehicle.

Team 8 conducted Data Farming experiments using their agent based model which represents situational awareness and the cognitive process to combine new sensor input with it to make identification decisions.

Team 9 used Pythagoras and a scenario developed for a prototype multi-agent system model of a civilian population to explore the response of the civilian population to insurgent, government, and stability force actions in a counterinsurgency environment.

Team 10 used the Joint Dynamic Allocation of Fires and Sensors (JDAFS) model which is being reviewed as a tool to support Joint Starting Condition data development. They explored a joint battlespace scenario in a Data Farming environment to identify possible improvements to JDAFS.

And finally, Team 11 built on research started in Canada on a systems dynamics model used to explore the use of non-lethal weapons in crowd confrontation situations. They used Data Farming and design of experiments approaches to help determine the most sensitive parameters and develop a robust set of rules of engagement

6 INVITATION

This paper has two purposes. The first is to describe the concept of Data Farming and give an overview of how it is being used worldwide. The second is to invite you to become part of our International Data Farming Community. We value openness, collaboration, and having fun in the process. By planting seeds of knowledge throughout the world we feel that we can grow the methods and tools to begin to provide answers to the difficult questions of our age. We invite you to contact us, we invite you to use our tools and methods, and we invite you to join us in person at our next International Data Farming Workshop in Mon-

terey, California, USA from 22 through 27 March 2009. There we will continue to strive to outline the landscapes of possibilities, discover surprises, and uncover those dynamic truths central to understanding questions that we share.

ACKNOWLEDGMENTS

Instantiating the idea of Data Farming in the real world of modeling and high performance computing has benefited from the patience, effort and interest of a large number of participants from around the world. The authors would like to acknowledge the integrity, audacity, and humility of all of the collaborators in the International Data Farming Community: the modelers, software developers, subject matter experts, analysts, administrative aids, students, teachers, and decision makers. We are especially grateful for the insight, interest, and time provided by men and women in military service around the world.

REFERENCES

Brandstein, A. 1999. *Operational Synthesis: Supporting the Maneuver Warrior*, briefing presented at the 2nd Annual Defense Planning & Analysis Society Symposium. Rosslyn, VA.

Henscheid, Z., Middleton, D. and Bitinas, E. 2007. Pythagoras: An Agent-Based Simulation Environment, *Scythe Issue 1*: 40-44. Monterey, CA.

Hoffman, F. and Horne, G. 1998. *Maneuver Warfare Science 1998*. United States Marine Corps Project Albert. Quantico, VA.

Horne, G. 1997. *Data Farming: A Meta-Technique for Research in the 21st Century*, briefing presented at the Naval War College. Newport, RI.

Horne, G. 1999. Maneuver Warfare Distillations: Essence Not Verisimilitude. *Proceedings of the 1999 Winter Simulation Conference*, eds. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 1147-1151. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc..

Horne, G. and Leonardi, M. 2001. *Maneuver Warfare Science 2001*. United States Marine Corps Project Albert. Quantico, VA.

Horne, G. and Johnson, S. 2002. *Maneuver Warfare Science 2002*. United States Marine Corps Project Albert. Quantico, VA.

Horne, G. and Johnson, S. 2003. *Maneuver Warfare Science 2003*. United States Marine Corps Project Albert. Quantico, VA.

Horne, G. and Meyer, T. 2004. Data Farming, briefing presented at the *2004 INFORMS National Meeting*. Denver, CO.

Horne, G. and Meyer, T. 2004. Data Farming: Discovering Surprise. *Proceedings of the 2004 Winter Simulation Conference*, eds. R. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 171-180. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Horne, G. and Meyer, T. 2005. Data Farming Architecture. *Proceedings of the 2005 Winter Simulation Conference*, eds. M. E. Kuhl, N. M. Steiger, F.B. Armstrong, and J. A. Joines, 1082-1087. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kleijnen, J., Sanchez S., Lucas T., and Cioppa T. 2005, A User's Guide to the Brave New World of Designing Simulation Experiments, *INFORMS Journal on Computing*, 17(3): 263-289.

Lampe, T., Schwarz, G., and Wagner, G. 2007. PAX: Designed for Peace Support Operations, *Scythe Issue 2*: 43-48. Monterey, CA.

Lauren, M., McIntosh, G., Galligan, D., Anderson, M., 2007. Recent Developments in the MANA (Map Aware Non-uniform Automata) Agent-based Model, *Scythe Issue 1*: 38-39. Monterey, CA.

Lawler, M. 2005. Data Farming Cultivates New Insights. *SIGNAL, International Journal of the Armed Forces Communications and Electronics Association, June 2005*. Fairfax, VA.

McDonald. M. 2008. *The Use of Agent-Based Modeling and Data Farming for Planning System of Systems Tests in Joint Environments*, briefing presented at the 76th Military Operations Research Society Symposium. New London, CT.

Meyer, C. and Davis, S. 2003. *It's Alive*. Ernst and Young Center for Business Innovation, Cambridge, MA.

Meyer, T. and Horne, G. 2007. *Scythe, Proceedings and Bulletin of the International Data Farming Community, Issue 1, Workshop 13*, The SEED Center for Data Farming, Monterey, CA.

Meyer, T. and Horne, G. 2007. *Scythe, Proceedings and Bulletin of the International Data Farming Community, Issue 2, Workshop 14*, The SEED Center for Data Farming, Monterey, CA.

Meyer, T. and Horne, G. 2008. *Scythe, Proceedings and Bulletin of the International Data Farming Community, Issue 3, Workshop 15*, The SEED Center for Data Farming, Monterey, CA.

Meyer, T. and Horne, G. 2008. *Scythe, Proceedings and Bulletin of the International Data Farming Community, Issue 4, Workshop 16*, The SEED Center for Data Farming, Monterey, CA.

Sanchez, S. and Lucas, T. 2002. Tutorial: Exploring the World of Agent-based Simulations: Simple models, Complex Analyses. *Proceedings of the 2002 Winter Simulation Conference*, eds. E. Yucesan, C. H. Chen, J. L. Snowdon, J. M. Charnes, 116-126. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schwierz, K. 2008. *Chronicle of Data Farming: The Story of Success*, briefing presented at IDFW 16. Monterey, CA.

Upton, S. 2004. *User's Guide: OldMcData, the Data Farmer, Version 1.0*. United States Marine Corps Project Albert. Quantico, VA.

AUTHOR BIOGRAPHIES

GARY E. HORNE is a research professor at the Naval Postgraduate School (NPS) in Monterey, California, USA. He holds a B.S in Mathematics and an M.Ed. in Education from the University of Maryland and a D.Sc. in Operations Research from the George Washington University. He is the originator of the Data Farming concept and the Executive Director of the SEED (Simulation Experiments and Efficient Designs) Center for Data Farming at NPS. His email address is gehorne@nps.edu.

KLAUS-PETER SCHWIERZ is a senior manager in the System Design Center of EADS DS (European Aeronautics Defense and Space Company, Defense & Security) in Germany. After education in Mathematics and Physics he holds a Diploma in Mathematics and a D.Sc. in Natural Sciences / Mathematics from the Albert Ludwigs University in Freiburg, Germany. His business position is characterized as a consultant for system and system of systems decisions with the help of simulation. He and many of his colleagues are members of the Data Farming Community, attracted by the methodology and are promoting the Data Farming Concept in Germany. His email address is Klaus-Peter.Schwierz@EADS.com.